

ASSEMBLY 1:

1 TTGGTCACTG CTCCTCTAAT TGTGTTGTCG ATGAGT GAGT TTGGGAGTT
51 CCTGTTCTT CTCR-GGTCG TGGGGGCCCT CACTGAGAIG TGTGAATAAC
101 CAGAGATUGA CAACTATOTG GTAGAGAACT TGGGCAAGCA CTTCTTACCT
151 TGGATGGACG GCGCTTCCTT GGACCACTTG AACCCAGCA TCTATGTGGG
201 CCTACGGCTC TCCAGTCTGC AGGCTGGTC CAGGAGAGAC CTCCTAGCTGC
251 ACAGCCTCAA GCTTGGTTAC CAGCAGTCC TCTACGGTC TGCGCTTCAGC
301 CAGGATGAGC GTGAGCTGCA GGGCAAGCTT TCCATCGCC AGCTCGCCCT
351 CTACCTGCTC GCPCTCAGAG CCAACCTGAA GTTGTCAGG GGCACACAAGG
401 GGGACAGGCT GGTCTCACAG CTCAAATGCT CCTGACGA TGAGAAGAGA
451 GCCATTGACA CAGCAGCCAT GCGAGGCTTG GCTTACCT GTCTGAAGCG
501 CTCAAACTTC AACCCCTGGTC GAGACAAAGG GATCAGCATG GCACTCAAGA
551 CAGTGCAGA GGAGATCTTG AAAGCCGAGA CCTCCGAGGG CCACTTGAG
601 AAIGTCTACA GCACCCATT GCGATTACAG ITCTCAIGA CTTCCCCAT
651 GCCTGGGCA GAACTGGGAA GAGCATGT TT GAGGGAGG GTTGCTTFC
701 TGCGGAGTCT GCAAGGATGGA GCTTCAAGA ATGCTCTCAT GATTTCGAG
751 CTGCTCGCG TCTGAAACCA CAAGACCTAC ATGAACTGTA TCTTCGGAGA
801 CTGTCTGGCA CCAACGAGTCA GTTGGGACG AGCTGCTGAG ACCATTCCTC
851 AGACGAAAGA GATCATCAGT GTCAGGCTG AGGTGCTTAG TCTCTGGCG
901 CGGTAAAGAC AGPCCATCTC GTTGTGGCG GGTGACACCG TGGAAATGTT
951 CCTGAAAGG GCGCATGAGT TAGGAGAATT CAATAGAA AAGCAGGTT
1001 CCTGTGTGAGG CCCCCTACTTA ACGTCTGTA TG-GGAAAGG GCGGGAGAA
1051 AGGGAGTCTT GCGAGCTCT CGGAGAAGA AACACCCAC TGTGCAAGG
1101 TAATGCTGAC IACAGACCAG AGGATGGA AAACATTGAG CTGAGGCTGG
1151 TTAGCTGTA CCGCTGTGAGC TCGCTCATCTT GAAACGCTC GCACTCGCC
1201 TAGGCTCTTA CCCCCTCTCC TGTGCTGCTT GAAACAGGAA CTCGCGCTGAC
1251 CCTGCTGCCA CCTGCTGTGAGC ACTTGTAGA ATGCTCTG GATCACCCTC
1301 AGGCAAGAGC CCTTCGAGGG CCGTAACTAATGCGGAGCT GGGAGCAGAG
1351 AGGCCAGCAT GTTCCCTGGAG AACATTTTGT GGAACAGCT GCGCAGCCTG
1401 GCGCTGAGG TCTCCCATGA AGGGCACCAG ATGCTCTGAT GGGCATGAAG
1451 GATCTGAGAC TCTTGGCAA AAAGGGAGA CGGGAAGGCG CAGGTGTTGT
1501 GAAAGACAGCT CGTTCTGTGG TTGGGCTGCTT GCAAGGAGG CTCCTGAGCC
1551 CGGGGGTAT GGCCTGACC CGAGCPCTC ACTGCTCTGT TAGAGTGGCA
1601 GCTCTGAGCT GTTGTGGCA CAGTACGCTCG GGAAGCTCA GCAAGGCTGC
1651 TCACTGCTG CCTCTGACCA AAATGAGCT TTGTGCTGCT GTGAAAAAAA
1701 AAAAAAAA AAAAAAAA AA

(SEQ ID NO:1

FEATURES:

5'UTR: 1 - 30
Start Codon: 31
Stop Codon: 1159
3'UTR: 1162

Homologous proteins:

Top 10 BLAST Hits

		Score	E
CFA 10800002456390	/altid=gi 1274777 /def=ref NP_009322.1 tr...	752	0.0
CFA 10800002456396	/altid=gi 293316 /def=jb AAE25526.1 transcri...	732	0.0
CFA 10800004926133	/altid=gi 339205 /def=gb AAAC057.1 (102648...	732	0.0
CFA 108000024049036	/altid=gi 1265167 /def=jb AAE31176.1 AAH01...	731	0.0
CFA 10800004926130	/altid=gi 4507109 /def=ref NP_00346.1 trans...	727	0.0
CFA 10800004926132	/altid=gi 339205 /def=gb AAAC056.1 (102647...	725	0.0
CFA 10800005170902	/altid=gi 7657659 /def=ref NP_06564.1 trans...	515	e-145
CFA 10800005218941	/altid=gi 4572454 /def=jb AAD23529.1 AF12128...	501	e-140
CFA 104000136745249	/altid=gi 11768124 /def=ref NP_071979.1 tr...	481	e-134
CFA 10800004926154	/altid=gi 4507407 /def=ref NP_001053.1 trans...	108	2e-22

EST:

gi 10725490 /dataset=dbest /taxon=96...	853	0.0
gi 10947399 /dataset=dbest /taxon=96...	846	0.0
gi 9121897 /dataset=dbest /taxon=9606...	846	0.0
gi 13280819 /dataset=dbest /taxon=96...	846	0.0
gi 13287907 /dataset=dbest /taxon=96...	833	0.0
gi 13286505 /dataset=dbest /taxon=96...	831	0.0
gi 8150776 /dataset=dbest /taxon=9606...	815	0.0
gi 5936410 /dataset=dbest /taxon=9606...	726	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi|10725490|adult adrenal gland
gi|10947399| mammary gland
gi|9121897| retinoblastoma
gi|13280819| adenocarcinoma cell line
gi|13287907| retinoblastoma
gi|13286505| embryonal carcinoma, cell line
gi|8150776| adult uterus
gi|5936410| adult uterus
gi|6888875| adult head_neck
gi|6888872| adult head_neck

Tissue Expression:

Human leukocyte

Isolate A:

1 CGAGGATTAA TCACTGACAG GAAGCTTCTT TTTCGAGC GGTGACCAAG
 61 TCGGCTTAAAT AAGGCTTCAG CAGGGCTAAT CCCARGAGTC TTTCGAGT
 101 CTTGCTTCACT GCTCACCCAG CTCCTGCTGTC CATGAGCCAC CTTGGGGCCT
 151 TCCTCTTCTT TCTGGGGCTC CTGGGGCTC TCACIGAGAT GTGTGAAATA
 201 CCAGAGATGG ACAGCCATCT GGTACAGAAAG TTGGGGCAGC ACCTCTTACG
 251 TTCGATGAG CCGCTTCTCC TGCACCACTT GAACCCAGC ATSTATGTGG
 301 CCCTACAGCT CTCCAGTCTG CACCTCTGGA CCAAGGAAGA CTCCTACCTG
 351 CACAGCCTCA TGCTTGGTTA CCACCACTGTC CTCC1ACGGT CTCCCTCAG
 401 CGAGGGATGTC BGTGACTGCC AGGCCAAGGC TTCCATGGC TAGCTGGCCC
 451 TCTACCTCTT CGCTCTCAGA GCGAACCTGGC ATGATCACAA TGGCCACCCC
 501 CACACTACGTT ACTACCACTA TGGCTCGGC ATTCTG1CCC TG1CTCTCCA
 551 CCAAAAGCAG GTCCATGACA BCGGIGGGGA CAAACTCTG TAACTGTGG
 601 AACCTTCA ACGAGGCCAC CATICTGTGG ACACAGTAGC TATGGCAGGC
 651 TTCTGCAATT ACGCTCTGAA CGCGCTCAAAC TCTCAACCTG TCTGGAGACA
 701 AGGAACTCA ATGGCCATCA GAAACAGTGGC AGAAGGAAATC TTGAAAGGCC
 751 AGAACTTCA GGGGCACTTT GGAAATGTCT ACACCAACCC ATTGGGATTA
 801 CAGCTCTCA AGACITCCCC CAI BCGTGGG GCAAAAGG GAAACAGCAAG
 851 TCTTAAAGGAG AGGGITGCTT TGTGGCCAG TCTGGAGAT CGAGCTTCC
 901 AGAACTCTT ATGATTTCG CACTGCPGC CGTTCGAA CTACAAGACG
 951 TAAATGAAAC TGATCTTCCC AGACITGTGG GCAACAGGAG TAACTGTGG
 1001 ACCAGCTGTT TAAACCATTC CTAGACACCA AGAAATCAGT AGTGTACCG
 1051 TCGAGCTGTT TAGTCTCTTG CGCGCTTACA GACAGTCAT CTCTGGTCTG
 1101 CGCGCTCA AGCTGGAAAGA TGTCTGAAAG AAGCCATG AETTAAGGAGG
 1151 ATTAACATAT TAAACACAGG TCGCTTTGTC AGGACCTTAA TAAACCTCG
 1201 TAAAGGGGAA AGCGGCCGGA GAAAGGGAGT TCTTGAAC TCTCGAGAC
 1251 TAAACACAGC TACTGTTGCA AGGTTATTGCT GACAAACA GAACTAAAGG
 1301 AAGAACATT TAGCTGAGGC TGGTAACTG GTCACCGCTG AGTTCGCTCA
 1351 TCTGACACT TCGCACACT CGCTAGGCTT CTACCTCCCTG TCTGATGTC
 1401 CTTGAAACG TAACTCGCCT TACCTCTGTC CGACCTCTG TCTGATGTC
 1451 GAAAGGCGTC TGGGATCAC CGACGGGACA AGCGCTTGA GGGCGCTATA
 1501 CAGGGCGTC TCTTGGAGCA GAGGGGAAAG CATTTGCTT GGGAGCTTT
 1551 TCTGGCCAGG TTGCGCTTGC AGGTTCGCA TGAAGGGCAC
 1601 CAGGGCTTCTT TATGGGCATG AGGTATCTCA GACCTCTTGA CAAAAACGG
 1651 AAGCTTCAAG TGGCAGGTGT TGTCAAGAGACC ACTGTTCTG TGTCTGGGT
 1701 CAGCTAAGAA AGCTCCTCA CCCGGGGGGC TATGGCCCTG AGTCAGCTC
 1751 TCACTTCTCTT TTTAGAGTG GCAGCTCCGA GCTCTGTC GCACTACTAGC
 1801 TGGGGAGACCTTAGCAGGGC TGCCTAGTGC CTGGCTCTGA CAAATTTAAA
 1851 GAACTCTGTTTGTGAAAAA TAAATTTAAA AAAAATAAAT AAAATAA

(SEQ ID NO.:

FEATURES:

5'UTR: 1 - 131
 Start Codon: 13
 Stop Codon: 13-2
 3'UTR: 13-5

Homologous proteins:

Top 10 BLAST hits

		Score	E
CRA 1080 00145367 6	/altid=gi 293316 /def=gb AAE15516.1 trans...	793	0.0
CRA 1080 00145370 0	/altid=gi 12742175 /def=ref NP_009222.1 trans...	793	0.0
CRA 1000 00145117 7	/altid=gi 339201 /def=gb AAE1051.1 (L03643...	792	0.0
CRA 1080 00145206 6	/altid=gi 12654675 /def=gb AAH017.1 AAH01...	792	0.0
CRA 100000191617 0	/altid=gi 4507409 /def=ref NP_000344.1 trans...	788	0.0
CRA 180000019513 2	/altid=gi 339201 /def=gb AAE1056.1 (L02647...	786	0.0
CRA 180000019092 2	/altid=gi 7657639 /def=ref NP_056564.1 trans...	561	e-159
CRA 16400013645249	/altid=gi 11968124 /def=ref NP_071079.1 trans...	554	e-156

CPA|180|0706|14941 /alt_id=gi|4573454 /def=jb|AAW8141|AF2124... 545 e-124
CPA|1800|074926134 /alt_id=gi|4507407 /def=refseq +1-53,11 tran... 128 1e-24

EST:

gi 10125490 /dataset=dbest /taxon=96...	858	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	835	0.0
gi 6818875 /dataset=dbest /taxon=9606...	726	0.0
gi 6818872 /dataset=dbest /taxon=9606...	726	0.0
gi 12158937 /dataset=dbest /taxon=960...	686	0.0
gi 10947399 /dataset=dbest /taxon=96...	680	0.0
gi 131287907 /dataset=dbest /taxon=96...	680	0.0
gi 9121897 /dataset=dbest /taxon=9606...	680	0.0
gi 10203819 /dataset=dbest /taxon=96...	680	0.0
gi 8150776 /dataset=dbest /taxon=960...	656	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi|10125490| adult adrenal gland
gi|5936410| adult uterus
gi|6818875| adult head_neck
gi|6818872| adult head_neck
gi|12158937| adult lung_tumor
gi|10947399| mammary gland
gi|131287907| retinoblastoma
gi|9121897| retinoblastoma
gi|10203819| adenocarcinoma cell line
gi|8150776|

Tissue Expression:

Human hippocampus

Sequence 1:
 1 MEEHIAAFEL LQWLGALTEM CEIPEMIRSHI VEPFAGHILP WMDFRISLPHI
 51 SMLVYVQRLR SFLQWTFKED LYIHLQFLKTY QV LIGRRAFS KIDWIDQKPE
 91 SPPAQLAVALM ALRANCEFPR GHGGRQLVLSQ LFWFLEDEFR AIIITAAMAGI
 131 APTVLEPRNIE NHGRPQRITM AIPTVREHIL KAQTPEGHFG NVYJTFIAQ
 171 FLMISFMRGA EIGTACLKAR VALLASLQDG AFQNALMISQ LLEPVINHKTY
 211 IIGLIFPPCOLA PPVMLEPAAE TIPQTEIIIS VTIQVLSLLP PYRQSIIVIA
 251 CNTVKEIVIHK AHLLGGFTYE QASLSGPYL TSMVGKAAGE REFWQLIQEP
 291 NTPLIQLQIAQ YPKDGETIE LELVSW
 (SEQ ID NO: 3)

FEATURES:

Functional domains and key regions:

PDOC00005 PS00005 PKC_PHOSPH_C_SITE
 Protein kinase C phosphorylation site
 Number of matches: 1

1	75-77	SLK
2	174-176	TVR

PDOC00006 PS00006 CK2_PHOSPH_C_SITE
 Casein kinase II phosphorylation site
 Number of matches: 6

1	67-70	TKED
2	90-93	SEED
3	174-177	TWFE
4	226-229	SLWD
5	249-252	TZID
6	302-305	STVE

PDOC00008 PS00008 MYRISTYL
 N-myristylation site

Number of matches: 7

1	12-13	GIVGAL
2	57-61	GLFLSS
3	86-91	GAAPSE
4	149-154	GAFTC
5	190-194	GMVYST
6	209-214	GAEIQT
7	230-235	GAFDNA

PDOC00009 PS00009 AMIDATION
 Amidation site

162-163 PGRP

SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	1.37	YES
max. Y	19	0.702	1.34	YES
max. S	5	0.974	1.38	YES
mean S	1-18	0.949	0.46	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

BLAST Alignment to Top Hit:

```
>CRA|108000024636216 /altid=gi|298316 /def=gb|AAE25526.1|
  transcobalamin II, TC II [human, endothelial cells,
  Peptide, 427 aa] . org=human /taxon=9606 /dataset=nraa
  /length=427
  Length = 427
```

Score = 732 bits (1870), Expect = 0.0
 Identities = 376/427 (88%), Positives = 376/427 (88%), Gaps = 51/427 (11%)
 Frame = +1

Query: 41 MRHISAFLEFLIGVIGALTEMCEIPEMICHILVEKLGQHILFWMPLNIEHIMPSIYVGLPL 210
Sbjct: 1 MRHISAFLEFLIGVIGALTEMCEIPEMICHILVEKLGQHILFWMPLNIEHIMPSIYVGLPL 60

Query: 111 SNIQAGTKEIDLHSLKLGYQQCLLGSASFSEDGICQGKFBMGMQLALYLIALRANCEFVR 340
Sbjct: 61 SNIQAGTKEIDLHSLKLGYQQCLLGSASFSEDGICQGKFBMGMQLALYLIALRANCEFVR 120

Query: 391 CHFEDRLVSQLWPLEDEKRAI----- 456
Sbjct: 121 CHFEDRLVSQLWPLEDEKRAI

Query: 457 -----DTAAMA GLAFTCLRSNFNPGRQRITMAIRTVREEILKAQTPEGHF 597
Sbjct: 181 YAVEPFHQGHHSVDTAAMA GLAFTCLRSNFNPGRQRITMAIRTVREEILKAQTPEGHF 240

Query: 598 GNVYSTPLALQFIMTSPMRGAELGTACLKAR'ALIASLQDGAFQNALMISQLLPVLNHKT 777
Sbjct: 241 GNVYSTPLALQFIMTSPMRGAELGTACLKAR'ALIASLQDGAFQNALMISQLLPVLNHKT 300

Query: 778 YIDLIEFPDCLAPVMLEPAETIPQTQEIIISVTLQVLSLPPYFQSISVLAGSTVEDVLK 957
Sbjct: 301 YIDLIEFPDCLAPVMLEPAETIPQTQEIIISVTLQVLSLPPYFQSISVLAGSTVEDVLK 360

Query: 958 KAHELGGFTYETQASLSGPYLTSMGKAAGEIEFWQLLFDPNTFLLQGIADYRPKDGETI 1137
Sbjct: 361 KAHELGGFTYETQASLSGPYLTSMGKAAGEIEFWQLLFDPNTFLLQGIADYRPKDGETI 420

Query: 1138 ELRLVSW 1158
ELRLVSW
Sbjct: 421 ELRLVSW 421
(SEQ ID NO:6)

HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	329.9	8.6e-246	2
CE00052	CE00052 lymphocyte_transmembrane protein KAP	3.2	2.9	1

Parsed for domains.

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11	[.	1	11	[.
PF01122	1/2	1	142	[.	1	143	[.
PF01122	2/2	143	376	.]	197	450	.]

Isocform_4:

1 MRHLGAFPLI LGVIAVATEM CRIPEMPSHL VEKLGQHILP WMDPLSLEHL
51 IPRHIVAHRLI SSIQQTTFKID IYHSLIMGY QQLLSSAFS KEDHPTQGHP
101 SMGQVLAINTL ALRANWHQHIC GHPTHTSYYQQ GLGILADCLH QMPVHRSTWD
151 KILYAVEEEFR QGHHLGVIATAA MAGLAFTVILK RSNENPGRRQ RITKAIRTVR
201 EELIKAQTPF GHPGNVYSTP LAIQFLMTSP MRGAEILGTAC LKARVALLAS
251 LQDGAFQNAL MISQOLPVIN HKTYIDLIFP DCIAPRVMLE PAAETIPQTQ
301 EISVTLAQL SLLPPYRQSI SVLAGSTVED VLKKAHELCC FTYHTQASLS
(SEQ ID NO:4)

FEATURES:

Functional domains and key regions:

PPDC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site
198-200 TVP

PPDC00006 PS00006 CK2_PHOSPHO_SITE
Caveolin kinase II phosphorylation site
Number of matches: 7

1	67-70	TKFD
2	90-93	SEID
3	147-150	SVVD
4	198-201	TVPE
5	250-253	SLQD
6	273-276	TYID
7	326-329	STVE

PDIC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 7

1	12-17	GVLCAL
2	57-62	GLFLSS
3	86-91	GSAFSE
4	173-178	GIAFTC
5	214-219	GNVYST
6	233-238	GAEILGT
7	254-259	GAFQNA

PDIC0009 PS00009 AMIDATION

Amidation site

186-189 PGFR

PDGC0426 PS00466 COBALAMIN_BINDING

Eukaryotic cobalamin-binding proteins signature
165-178 SVITAAAMAGLAFTC

SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

BLAST Alignment to Top Hit:

>CFA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|
transcobalamin II, TC II [human, endothelial cells,
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa
/length=427
Length = 427

Score = 754 bits (2026), Expect = 0.1
 Identities = 399/427 (93%), Positives = 399/427 (93%), Gaps = 27/427 (%)

Query: 1 MPHICIAFLPFLGVLGALITEM'EIPEMDSHLVEKLQHQHLLPWNMDRLSLEHLNPSIYVGLRL 60
 Sbjct: 1 MRHICIAFLPFLGVLGALITEM'EIPEMDSHLVEKLQHQHLLPWNMDRLSLEHLNPSIYVGLRL 60
 Query: 61 SSLQAGTKEDLYLHSLMIDYQQCLLGSAFSEDDGDCQGKP.SMGQLAIYLLALRAN---- 115
 Sbjct: 61 SSLQAGTKEDLYLHSL IDYQQCLLGSAFSEDDGDCQGKP.SMGQLAIYLLALRAN
 Query: 116 -----W-----HDHKGHPHTSYYQYGLGILALCLH)KRVHDSVVDKLL 153
 Sbjct: 121 SHKGDRLVSQLKWFLEDEKRATGHDHKGHPHTSYYQYGLGILALCLH)KRVHDSVVDKLL 180
 Query: 154 (AVEPFH(GHHSVDTAAMA GLAFT CLKRSNFNPGR)QRIT:MAIRTVEREILKAQTPEGHF 213
 Sbjct: 181 (AVEPFH(GHHSVDTAAMA GLAFT CLKRSNFNPGR)QRIT:MAIRTVEREILKAQTPEGHF 240
 Query: 214 GNVYSTPLALQFIMTSPMRGAELGTACLKARVALIASLQDGAFQNAIMISQLLPVLNHKT 273
 Sbjct: 241 GNVYSTPLALQFIMTSPMRGAELGTACLKARVALIASLQDGAFQNAIMISQLLPVLNHKT 300
 Query: 274 YIDLIFPCLAEFVMLEPAAAETIPQTDEIISVTLQNLSSL:FYFQSIYLAGSTVEDVLK 333
 Sbjct: 301 YIDLIFPCLAEFVMLEPAAAETIPQTDEIISVTLQNLSSL:FYFQSIYLAGSTVEDVLK 360
 Query: 334 KAHELGGFTYETQASLSGPVLTSMGKAAGEREFWQLLRDPNTFLQGIAADYRPKDGETI 393
 Sbjct: 361 KAHELGGFTYETQASLSGPVLTSMGKAAGEREFWQLLRDPNTFLQGIAADYRPKDGETI 420
 Query: 394 ELFLVSW 400
 Sbjct: 421 ELFLVSW 427
 SEQ ID NO:()

HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	90e.3	8.6e-269	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	1.2	2.9	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11	1	11	1.2	2.9
PF01122	1/2	1	115	1	115	241.3	1.4e-68
PF01122	2/2	117	400	145	450	660.5	8.7e-195

1 ATATGATGCG GAAATATGTT GTTTTCTAT TTTTATGTT GCAACCTTA
51 GCTTGACTTC CAGCTAGGTA GGCAGGGGG TGTGTTGTT TTTTACTTGG
101 ACGCCCTAAAT TAACTTGGG ATGAAAGGIG AGTTTCTGG AGGATGTTG
151 GATTTGTTGAGCTGAGCTGA CGATTAACGGA KTTATCTGTT CAT GCCTA
201 TGTCTAGGCTT TCTG GAAATGGGA TCTAAATAT TCT ACTGA
251 TTTTAAAGGAA AGCTGGGG AGTAAAGA TCTGAAAGA GCA AAGT
301 AGT ATTCG CAGCTCTTC AGTGTGAGT CTCAGGTTG CCAAGGTTG
351 AGT TGG AGCTTGGCT CTCAGGTTG CTCAGGTTG CTCAGGTTG
401 TCGA CGTG CTC TCTAG TGTGAGCTG AGGTTTGT CACT GCGA
451 TCTAT ATTTTAACTTTCAGTAAAGTCTG TAAAGAGAA GCTCTCTG
501 TCTCTTCTGAGGCTG TGTGAGCTG ATAGTGTG CTCAGGTTG
551 TTCTTAAAGA AGCTGGGG AGTAAAGA TCTGAAAGA GCA AAGT
601 TACAGGTTG CTCAGGTTG CTCAGGTTG TAAAGAGAA GCTCTCTG
651 ATATGTTAG CTGAGCTGAGCTG TAAAGAGAA GCTCTCTG
701 AGGTTAGA GAAAGAA TGTAAAGA TGTAGGTTAA GCGAGCTGG
751 TAAAGTGTGAA AAGTGTGAGCTG TAAAGAGAA GCTCTCTG
801 AGGTTAGTGTGAGCTG TAAAGAGAA GCTCTCTG
851 TGTGTTAGG AGTCTTGG CTCAGGTTG CTCAGGTTG
901 AGTTTCTGTTGAGCTG TAAAGAGAA GCTCTCTG
951 TAAAGAGAA GCTCTCTG TAAAGAGAA GCTCTCTG
1001 AGTATTTAGG TAAAGAGAA GCTCTCTG
1051 ATGTTAGG TAAAGAGAA GCTCTCTG
1101 ATGTTAGG TAAAGAGAA GCTCTCTG
1151 ATGTTAGG TAAAGAGAA GCTCTCTG
1201 TGTGTTAGG TAAAGAGAA GCTCTCTG
1251 TAAAGAGAA GCTCTCTG
1301 AGTATTTAGG TAAAGAGAA GCTCTCTG
1351 AGTATTTAGG TAAAGAGAA GCTCTCTG
1401 TAAAGAGAA GCTCTCTG
1451 AGTATTTAGG TAAAGAGAA GCTCTCTG
1501 AGTATTTAGG TAAAGAGAA GCTCTCTG
1551 AGTATTTAGG TAAAGAGAA GCTCTCTG
1601 AGTATTTAGG TAAAGAGAA GCTCTCTG
1651 AGTATTTAGG TAAAGAGAA GCTCTCTG
1701 AGTATTTAGG TAAAGAGAA GCTCTCTG
1751 AGTATTTAGG TAAAGAGAA GCTCTCTG
1801 AGTATTTAGG TAAAGAGAA GCTCTCTG
1851 AGTATTTAGG TAAAGAGAA GCTCTCTG
1901 AGTATTTAGG TAAAGAGAA GCTCTCTG
1951 AGTATTTAGG TAAAGAGAA GCTCTCTG
2001 AGTATTTAGG TAAAGAGAA GCTCTCTG
2051 AGTATTTAGG TAAAGAGAA GCTCTCTG
2101 AGTATTTAGG TAAAGAGAA GCTCTCTG
2151 AGTATTTAGG TAAAGAGAA GCTCTCTG
2201 AGTATTTAGG TAAAGAGAA GCTCTCTG
2251 AGTATTTAGG TAAAGAGAA GCTCTCTG
2301 AGTATTTAGG TAAAGAGAA GCTCTCTG
2351 AGTATTTAGG TAAAGAGAA GCTCTCTG
2401 AGTATTTAGG TAAAGAGAA GCTCTCTG
2451 AGTATTTAGG TAAAGAGAA GCTCTCTG
2501 AGTATTTAGG TAAAGAGAA GCTCTCTG
2551 AGTATTTAGG TAAAGAGAA GCTCTCTG
2601 AGTATTTAGG TAAAGAGAA GCTCTCTG
2651 AGTATTTAGG TAAAGAGAA GCTCTCTG
2701 AGTATTTAGG TAAAGAGAA GCTCTCTG
2751 AGTATTTAGG TAAAGAGAA GCTCTCTG
2801 AGTATTTAGG TAAAGAGAA GCTCTCTG
2851 AGTATTTAGG TAAAGAGAA GCTCTCTG
2901 AGTATTTAGG TAAAGAGAA GCTCTCTG
2951 AGTATTTAGG TAAAGAGAA GCTCTCTG
3001 AGTATTTAGG TAAAGAGAA GCTCTCTG
3051 AGTATTTAGG TAAAGAGAA GCTCTCTG
3101 AGTATTTAGG TAAAGAGAA GCTCTCTG
3151 AGTATTTAGG TAAAGAGAA GCTCTCTG
3201 AGTATTTAGG TAAAGAGAA GCTCTCTG
3251 AGTATTTAGG TAAAGAGAA GCTCTCTG
3301 AGTATTTAGG TAAAGAGAA GCTCTCTG
3351 AGTATTTAGG TAAAGAGAA GCTCTCTG
3401 AGTATTTAGG TAAAGAGAA GCTCTCTG
3451 AGTATTTAGG TAAAGAGAA GCTCTCTG
3501 AGTATTTAGG TAAAGAGAA GCTCTCTG

3851 GACGTTTGTG CCAATTAACCT GCAAGGCAAG GCGGCGAAT GCGTTGGGGT
3861 CTGGGGGATG TT TAACTGGG CTTTGGGGGTG CAAATATCTG CAACTGAGGCA
3871 CTTCAATTTAA CG TAACTGAA TTTTTTTATG TCTAAAGTG CTTAAACAACTG
3881 CTGTTTTCG TT TCTCTCG CTTAGTAGCG AGGTAAGGATG CAACTAAACG
3891 TGTCTTCA GTTTAAGCT TATATGAT GTCAGGAT TGAGCTAGAT
3901 CGCCTCGAA ACCTATATGC AACACGGGCT CTTTGTCTG CTTCCTTCT
3911 TCTCTTCTT TT TTTTTTTT TTTTTTTCTC TCTACAGCTG AGCTTCTCT
3921 ATATGCTCA GGTCTGTTT TCTATTCCTA GCGCAGCTG ATGCTCTTAC
3931 CTGCGCTCC CAACTCTG CAACTACAGG CTTTACCCAC AGCTCTTCG
4001 TCTCTTCTP GTAAATGCTCA AGCTAAATCA GCTCTTACT GAGGGCTG
4051 GCTTGGAGG CAACTACCTC AGCAAGGT GCTTAACTA CTCCTCTCA
4101 ATCTGGCTG TCTCTTCTT TCTATAGCT TCTCTGCG AGCTCTCT
4151 CTCTCTTAC AGCTCTACCT CAACTCTP TCTCTTACCT TCTCTTACAC
4201 TCACTGGCT GGAATGAGA GCGCTGGGG AGCTCAGGGA ACCTCTCTC
4251 CTCTTCTG AATGAGGAA AGAAAGAAAGA GGTGACGGGAGG AGGCTAAAG
4301 AATGAGGAA CAACTCTA GAAATACGAA GACACACCA TAAAGCTTA
4351 TAACTTCTG GAACTGGAA TAAATAGGAA GAAAGGAGG AGAAAGCTT
4401 TCTCTTCTG AGAAAGGGAA AAAAATGAGC CCGCTCTG TCTATACAGC
4451 TGGGAGAAA GCTCTTCTG AGCTCTCTG GGGGCTCTAC AGAAAGCTT
4501 TATCTGGCT TCTCTGCTG TCTCTGCTA CTTTAACTCA AGCTCTCT
4551 GGTCTCTG GGAATGAGA TAAATAGGAA TCTGAGCTG AGCTCTCT
4601 AGCTCTCTG TCTCTGCTG TCTCTGCTG TGGGAGCTG AGAAAGCTT
4651 GCACTTAA GCTCTTCTG AGCTCTCTG AGCTCTCTG TAAAGCTTA
4701 GCACTTAA CA GGAATGAGA AGCTCTCTG AGCTCTCTG TAAAGCTTA
4751 TAACTTCTG TCTCTGCTG TCTCTGCTG TGGGAGCTG AGAAAGCTT
4801 GGTCTCTG AGCTCTCTG TCTCTGCTG AGCTCTCTG TGGGAGCTG
4851 GATGCTAA GGTAAAGGGG TCTTAACTCTG GCTCTTCTG TAAAGCTTA
4901 TCTCTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG AGCTCTCT
4951 AGCTCTCTG AGCTCTCTG TCTCTGCTG TGGGAGCTG AGAAAGCTT
5001 GGTCTCTG TCTCTGCTG TCTCTGCTG TGGGAGCTG AGAAAGCTT
5051 AAAAGGCTG AGCTCTCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5101 GGTAAAGGG AGCTCTCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5151 ATATGAGGAA TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5201 GGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5251 GGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5301 GGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5351 AAAAGGCTG AGCTCTCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5401 GGTAAAGGG AGCTCTCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5451 ATATGAGGAA TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5501 GGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5551 TAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5601 GGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5651 GAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5701 AATGAGGAA GAACTTCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5751 TAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5801 TAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5851 TAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5901 TAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
5951 TAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6001 TAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6051 TAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6101 AGCTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6151 AGCTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6201 TGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6251 TGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6301 TGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6351 TGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6401 TGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6451 TGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6501 GAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6551 AGCTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6601 GGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6651 GAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6701 TGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6751 GAACTTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6801 AGCTCTCTG AGCTCTCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6851 GGTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6901 AGCTCTCTG TCTCTGCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
6951 AATGAGGAA GAACTTCTG TCTCTGCTG AGCTCTCTG TCTCTTCTG
7001 ACTTTGGGAG GCGAAAG TAG GTAGATCTCA TG GGTCTAGG AGCTCTGAG
7051 CAGTCCTGGCC AACATACTGAA AACCTCTGCTG CTACTGAAAG TACAAAAAAT

651 TATGAAAGG TGTGTTGCG CAA TGTAT CTTCTTCTT TGGGAGGCG
 661 ACGAGAGA ATGGTTGAA CTCGGAGGC GEMTTGGA GTGAGCTGA
 671 AGCATCTAG TGCCTGAG CTCGGAA AAG AAGAGAAA CTCCTGCTGA
 681 AGGAAATTA ATAAATAAA TAAATA ATA AAGAGGAG CCGGATATGG
 691 TGAATATG CTCATAGG TGGCGGGG AAGATGATG CTCGACAGG
 701 ACCCTTCA GCTTCTT GTCCTTCTT GTCATCTT ATTAATATC
 711 TCTGGCTAC AAT AAGAAGTATCA ATTCACCTT TTTTGAGCT
 721 TTATCTAG TTTCTGAG TTAGAATT TT CCGCTT CTGATACGG
 731 TGGGTGGCT TGTGAGCA GCGTCTGG CTGGAGTTC CTGAGAAGG
 741 CATTAA TGG CTTTCTCA GCTTCTT CTCAGAGAT GACGGGAGT
 751 GCGAGCTCA GCTTCTAG GAGTATGG CTCATACCT GTCGGCTTC
 761 AGAGGCTACT CTGTTCTT AAGCTTAC AAGGGGACA GCGTCTTC
 771 AGCTT AAA TTTCTTG AGATAGAA GAGGCGATC GCGAGCTCA
 781 TACCATCTTC TACGGGGG GAGTATGG GAGGCGCTAT GAGATATAT
 791 TCTGAGCTCA GAGTAAAC TTTCTTCTT GTCCTTCTT ATTCCTCA
 801 TCTGAGCTTC TTTCTTCTT GAGTATGG GAGGCGCTAT GAGATATAT
 811 TATTTCAT AAGCTTGGG TTTCTTCTT TGGGAGG CCGCTCTAA
 821 TCTGCTCT GAGTATGG GAGTATGG GAGGCGCTAT GAGATATAT
 831 TACAGGATG AATCTCTG TTAAGGAGC TACAGGAGT GAGATATGG
 841 AGCCAGAG AATCTCTG TACAGGAGC TGGGAGAA GAGGCGCTAT
 851 GCTTATAT TTTCTTCTT GAGTATGG TGGGAGG GAGGCGCTAT
 861 TGACCCAGCA TA AACAGG AAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 871 TGTGCTGAG TTTCTTCTT GAGTATGG GAGGCGCTAT GAGGCGCTAT
 881 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 891 TCCCTCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 901 ATAACGAGCTT TTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 911 TGTGCTGAG TTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 921 GCGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 931 AGGCAATAT GAGTATGG TGGGAGG GAGGCGCTAT GAGGCGCTAT
 941 GAGCAATGG TGGGAGG GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 951 AGGTCCTGAG TTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 961 GCACTTCTTC GAGTATGG TGGGAGG GAGGCGCTAT GAGGCGCTAT
 971 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 981 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 991 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 1001 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 1011 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 1021 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 1031 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 1041 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 1051 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT
 1061 AGGCTGCTCA GCTTCTTCTT GAGGCGCTAT GAGGCGCTAT GAGGCGCTAT

10651 CAGTGTGGCG ACGACACAG TAAATGGAGA TCCCTGGGAA TATGCGTTCT
 10701 GTGATAACCA CTAAAGAAAT GTTGGGATCG ATGGATGGAT GGATGGATGG
 10751 ATGGATGGAT CGATGGATCG ATGAGGAGAG ACACATTTG GTTAAATTA
 10801 ATACACATG ATAAGCCTCA ATGAGGAGAT CATCCAGGCT TCTCTGAGA
 10851 CAGGGCTGCA GAGATTTAT TGGATTTG AATTAAGAAAT CGAGGAGAG
 10901 CAGATTCGA CGCACTCATG TCTTAAATCG CAAACCTTGT GGAGCTGGAG
 10951 CGCGGTTGGCT CACGTGAGGT CAGGAGTTC AGACACGGCT GGCTAAATG
 11001 CGAAACTCG ATCTTTAAA AATACAAAAA ATGAACTCGG TGCTGTGGG
 11051 ATGCTGTGA ACCGAGCGA AGCTACTCG GAGGTGAGG CAAGAGAATC
 11101 ATCTGACCTG CAGGATGGA GTTGTGAGT ACGCAAGAAT ATGCGATTC
 11151 ACTCGCTCT CGGTGAGAGA STAAGACATG GTCTAAAAA AAAAAAA
 11201 AAATGAGAGA GAGGAGGTT CGAGATGTT GGTGCTGCTT AATACCTAG
 11251 CACCTGGG CGCTGAGCA GATGGATTC CGAGGAGAG GAGTGGAGA
 11301 CGAGCTGGG CAAACATGGT AAACCTGTT TTTACAAAAA CAGGAAAGAT
 11351 TACGAAAGCA TGGTGGGAGA CACCTATATC CGAGGAGATG AGGGAGGCTG
 11401 AGCCACAAAGA ATCACTTGA CCGGGGAGA AGAGGPTGCA GTGAGGCGAG
 11451 ATGCGCCAC TGCACCTCG CCGGGGAGA AGGGGAGAG TGGGCTGAGA
 11501 GAAAACAA CGATGGATAG ATGGGAGCGA AGAAGAGGCA GAAAGAACAA
 11551 AGAAGAGAG GTCACAGAG TTTGGGGGAA TTTGGAGGAA TGGGCTGAGA
 11601 AAAGGTTGGG ATCTGGGAGA ATGAGTGGGAG GTCAGAGAAT GAGGAA
 11651 GAGAGCTGA CGCCATGAG CAAACAGAG TGGGTTTA ACAAATGAA
 11701 GGTCTAGAA CGGGCTCTG GAGTGGCTAGG TGCGAAAGA TGGGGAGAA
 11751 GAAAGCTGGA AGGAGAGAA CGAGTGGGGG GCGCTGATTT TGGGNGGGGG
 11801 GAAAGGAGGA CGGGGGAGC CGGGGATGGAA CGGAGGTTT TGGGAGGAG
 11851 GGAAGGGGGG CGGGGGGAGT GAGGCGAGGT TTAGTTGAGG GAGGTTGGG
 11901 TCGATTGAT GGAATGAA AGATGGGAGA GGAACAGTGT CAGGTTGCTT
 11951 TGAAGGAAAGAAGAAGAAGAATGAACTGTTCTTCTTCTTCTTCTTCTTCTT
 12001 CTGTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12051 ATCGATCTCTC AGCGCGAGA GATGAGATGT GTGAGCTTCTT AATGTTGAG
 12101 TCTTTGCGG CCGTAAAGAG AGGGCTATCT TGTGCTGTTG GGGGAGAG
 12151 TGGGATGT CGGGAAAGG CGGGGATGAGT TAAAGGTTT TAACTGGAGA
 12201 TCGACCTG CGATGTTGAG CCGGGGAAAGGCTTCTTCTTCTTCTTCTTCTT
 12251 GCGAGGAAAGAAGGAGGAAAGAAGGAGGAAAGGCTTCTTCTTCTTCTTCTTCTT
 12301 TGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12351 TATGAAACAGCTGAGATA CGGGGAGGTT TGGGAGGAGAG GGGGAGGAG
 12401 TGGAGGAAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
 12451 TGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12501 AGGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12551 TGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12601 TGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12651 TGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12701 TGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12751 TGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12801 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12851 TGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12901 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 12951 TGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13001 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13051 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13101 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13151 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13201 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13251 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13301 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13351 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13401 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13451 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13501 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13551 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13601 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13651 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13701 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13751 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13801 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13851 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13901 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 13951 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 14001 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 14051 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 14101 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 14151 GGGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT

1931 TTTTCTTGTG GATGGAATTG TGTCTATGTT GCTTAACTTA AGCTTAAAT
 1941 GATGTTTTT AATGTA TGG AAATTTAA TTTTTTTT GATTTTTT
 1951 GTCGTTTA GCTTGGAGT AG-TGCATT A-TTGTATGA AT-TAATATA
 1961 TT-TTAAAT TT-TTCTTTT AT-TAATGTT GATTTTTT AT-TTGTATTA
 1971 GGTGTTTT GAAATTTAA CTTAACATGA TCTTTTTT TCTTTTTT
 1981 AATGTTATGTTAAGGTT TTGAGTTAT TT-TTGTATTA GATTTTTT
 1991 GAGTTTAAG AGGTTTGTG TGGTGTGGG AT-TAGAGGG AGTTTAA
 2001 TT-TAAATTTA AA-TTGTAA GCTTAAATGG GGTGTTATGTT GATTTTTA
 2011 GT-TCTTGGG TGGCAGGGGG AGAAAGTAA CT-TTAAATCT TT-TTAACTA
 2021 GATTTTTA AT-TTGTATTA AT-TAGAGGG AGTTTAACT AT-TTAACTA
 2031 GAGAAAGGG GAAATGTTATG GAG-CATAGA GAAATGTTG TGTGTTGAGG
 2041 TGT-TTGTG TGTCTTGGG ATGAGCTGGG CTGTTATGTT AT-TAATGTT
 2051 CAG-CCTTCT CGTGTGATAT TGTACACATT CTAGGTACA CGATGAGTT
 2061 CTGATTCTA TGTCTTCTAA GAT-TAGGCCCC CGAGAGGGG GAAATGTT
 2071 GAG-TGATTA CATAATAGAT GCGCTTTTC TGGAGGTTAT GGTATGTT
 2081 AT-TGTTGAGG GGTGAGGGGG TGAAGGGGG GGTGAGGGGG TTGAGGGGG
 2091 AAGGAGGTTG AAGTGTGAGAT CACAGAGCTG TGGTCAAGTG GCTTGGGAAAG
 2101 GGTITGAGGA GTGIGGGGCC AAGAAATTTG GAAGGGATTG TGCTTCTGTT
 2111 GGTGAGGACT GCGCTCTCCC TTATGGACAA CAGGGAAATTC TTTTTT TCCC
 2121 AT-TGCTTATG CTCTCTGTA GATATGAAAC ACACTTCTG TGTGAGGG
 2131 CCTACTTAACT CTCGGTGTG GGGAAAGGGG CGGAGAAAG CGACTTCTG
 2141 CAGTTTCTG GAGACCCCAA CACCTTCTG TTGCTTCTG ACTTATGGG
 2151 TGAATCTG GATCTGTCG CTACCCAAAG CTTACTCTG CAATAGGTT
 2161 CATAAATCTA CCCCAGCTT CCTCTAGACC CTGCTGGGCC ACACTTCTAC
 2171 AAAATCCTG ATGCTCAAAG TTGATATATA TATATTGAAAC TGAACCCCTTA
 2181 GCAITTTAT GCAAGTTACT GTGAAATTG TAGGAAACCA GAAAGATTAC
 2191 AAAAAAAA AAAAATAGA AGAAAATTA CATTTCTAG GATATACTAC
 2201 CTAGAATAA CGTCTTTAT TTTGAGATGG AGTTTCGTC TTGTTGCCCC
 2211 GGTGAGGTTG CAGGGTATG ATCTGGCTG GCTGCAACT CCGCTCTCTG
 2221 GGTICATGTC ATTCTTCCAC CTCGGCTTC CTAGAGGCCA AGTGGTCTGC
 2231 CTGGCTTCTG CTCCAAAAGT TCTGGGATTA CAGGCATGAG GCAAGGCC
 2241 CAGGAAAAT TACTTAACTT TTCTTCTAGA TACTTTTAA AAATATGGCA
 2251 GTAAGTTTTT CATAAAAAT GAGGGATGC TATGCACTGG AAATTTAATG
 2261 TTGGCCACAT GTAACTTA AAAATTCAT ATAIGTGTAT ACATATATAT
 2271 GAAATATATA TATACACACA CACATATATA TGTATACATA TATATACACA
 2281 TATATATGTA TACATATATA CACACATATA TGTATACATA TATATACACA
 2291 CATACTACA TATATACACA CACATACATA TATACACACA TATATATACA
 2301 CACATATATA CACACATGCA CACATATATA TGTATACATA TATACACACA
 2311 TGTATATGTA TATATACACA CACATATACA TACATATACA TACATACACA
 2321 TACACATACA CACACACATA TATACACACA TATATACACA TATATACACA
 2331 CACATATATA TGATACATA TATATACACA CACATATACA TACATGTTATA
 2341 TACATATATA CACATATACA CACATACACA CACATATACA TACATGTTATA
 2351 CACATATATA CACACATGTA TACATATGTA TACACACACA TATATGTTATA
 2361 CATACTACA CACATACATA TGTGTACATA TATACACACA TACATATGTA
 2371 TACATATATA CACACAT

(SEQ ID NO:5)

Reported:

FEATURES:

Exon: 3'31-3034
 Intron: 2934-5164
 Exon: 5535-5762
 Intron: 5763-7871
 Exon: 7872-7741
 Intron: 7742-10000
 Exon: 10001-10173
 Intron: 10174-10298
 Exon: 10299-10485
 Intron: 10486-12027
 Exon: 12028-12193
 Intron: 12194-25821
 Exon: 25822-25939

Allelic Variants (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
921	C	T	Beyond ORF(5')			
1781	C	T	Beyond ORF(5')			
1890	G	A	Beyond ORF(5')			
3839	A	G	Intron			
3910	G	A	Intron			
6671	G	A	Intron			
6945	-	A	Intron			
6952	A	T	Intron			
7457	G	A	Intron			
7870	T	A	Intron			
8089	T	C	Intron			
8111	C	T	Intron			
9169	G	C	Intron			
9362	C	T	Intron			
9782	G	T	Intron			
10493	G	A, T	Intron			
11180	A	G	Intron			
13086	T	C	Intron			
13183	T	C	Intron			
21240	C	G	Intron			
21695	A	G	Intron			
21698	C	T	Intron			
22433	C	A	Intron			
22446	C	-	Intron			
23378	C	T	Intron			
23642	A	-, T	Intron			
23744	T	C	Intron			
23873	A	-	Intron			
24764	G	T	Intron			
24919	T	C	Intron			
24945	G	A	Intron			
25092	C	T	Intron			
25421	T	G	Intron			
25441	C	T	Intron			
25664	C	T	Intron			
26161	A	-	Beyond ORF(3')			

Context:

DNA

Position

921 TTGGAGATATTTAAGGTATAAGTGCTTCACAAATTGAGCTGAAAGGGAACTGTTAGGA
 TGATCTTGCCTAACCTCTCATCTCACACAGGAAGAACATATTTAACTCGAGAGGTTAA
 GTGACCTCGC:AAAGTCACACAGCCACCACTAGTTAACCTCGTATACTTGTATTCTCCTGT
 GGGGCTGGGCAGATGAGGAATCTTTGTTCTCTTCCCTGTTGCAGAGATTTTTGAG
 GTTACTTTCCSAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTGTCTCGATTCAT
 {C, T}
 TCATTCTTTTATTTTATTTTGAGACAGGGCTCACTTTGTCAACCCAAAGCTGGA
 GTGCAGTGGTGAATCTGGCTCACTGTAGCCTCCACCTTGGTTCAAGCGATCCTCC

	[1, 3]	GCTTCAAGGTTGGTGTGAGATGCGGAGCTTAATTTCGTATTTTAGTAGAGACAGGGTTT TACATATGTTGCGAACTTGGTTTGTACGAACTGGCTGTCATCGCGCTGCTGGTGGGTT CCGGAGTGTGTTGAAAGGGATGGGATGCGTGCGGATGCGGCGGATGTTATTTTTRAGGCA AGGTTCTCTCTGTTCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT CAAACTATCCCTAGTACGGACTTACCTTACCAACACACATCTAATTTTTTT
21240		TGAACTGAACTTAAAGCTCTCCGCGGCTGAAAGGACCTTCAACCTTAACTCTGCA GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTGGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA TGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA [1, 3]
		CTAAAGCCTCTTATAAATACACCTTGGCTGTTAACTGAACTGAAATGGGCG ACAGAGACATTTAGAATAAAACGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA TGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA
21695		CATTTGAACTTAAAGCTGAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA TGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA [1, 3]
		CTAACTGAACTTAAAGCTTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA TGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA
22058		CGCAAGTGTGAACTTAAAGCTTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA [1, 3]
		CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA
22233		CGCTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA [1, 3]
		CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA
22245		CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA [1, 3]
		CGAGCTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA
22375		ACAAACCGCTTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA AACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA CTGGCTGCTGGCTGAAAGTTCAGCTGCAATTTAGCTCACTGCGCT GAACTTGCGGTTAACTGGCTGCGGAAACCGCTGCGGAAAGGAAATTTTGTATGAGA

FEATURES:

Exon: 2157-2179
 Intron: 2144-2149
 Exon: 5871-5883
 Intron: 5884-5892
 Exon: 7613-7721
 Intron: 7722-8143
 Exon: 8150-8361
 Intron: 8363-10101
 Exon: 10102-10274
 Intron: 10275-10349
 Exon: 10400-10586
 Intron: 10587-11129
 Exon: 12129-13294
 Intron: 12295-25922
 Exon: 25923-26040

Allelic Variants (SNPs):

Position	DNA			Protein		
	Major	Minor	Domain	Position	Major	Minor
1 2	C	T	Beyond ORF(5')			
1-42	C	T	Beyond ORF(5')			
1 41	G	A	Beyond ORF(5')			
2 40	A	G	Intron			
3 41	G	A	Intron			
5 32	G	A	Intron			
7 43	G	A	Intron			
7 41	T	A	Intron			
8 40	T	C	Intron			
8 42	C	T	Intron			
8 40	G	C	Intron			
9 43	C	T	Intron			
9 43	G	T	Intron			
10 34	G	A T	Intron			
10 61	A	G	Intron			
11 187	T	C	Intron			
11 184	T	C	Intron			
11 181	C	G	Intron			
21 96	A	G	Intron			
22 159	C	T	Intron			
22 134	C	A	Intron			
22 346	C	-	Intron			
22 416	C	T	Intron			
22 143	A	- T	Intron			
22 445	C	C	Intron			
23 974	A	-	Intron			
24 865	G	T	Intron			
25 140	T	C	Intron			
25 246	G	A	Intron			
25 193	C	T	Intron			
25 319	T	G	Intron			
25 614	C	T	Intron			
25 755	C	T	Intron			
26 066	A	-	Beyond ORF(3')			

Context:

DNA

Position

1022 TTGGAGATATTTAAGGTATACTGTCTTCACAAATTGAGCTGAAAGGGAACTGTTAGGA
 TGATCTTGCCTAACCTCTCATCTCACACAGGAAGAACTATTTAAACTCGAGAGGTTAA
 GTGACCTGGCCAAGTCACACAGCCACCACTAGTTAACTCGTATACTGATTCTCCGTGT
 GGGGCTGGGAGATGAGGAATCTTTGTCTCTCCCTGTTGCAGAGATTTTTGAG
 GTTACTTTCCGAGTTCTGGCAAGTACCCCTGCTCTGGTAGCTTGTCTCGATTCAAT
 [C,T]
 TCATTCTTTTATTTTATTTGAGACAGGGTCTCACTTGTCAACCAAGGTGGA

FIGURE 3, page 16 of 22

